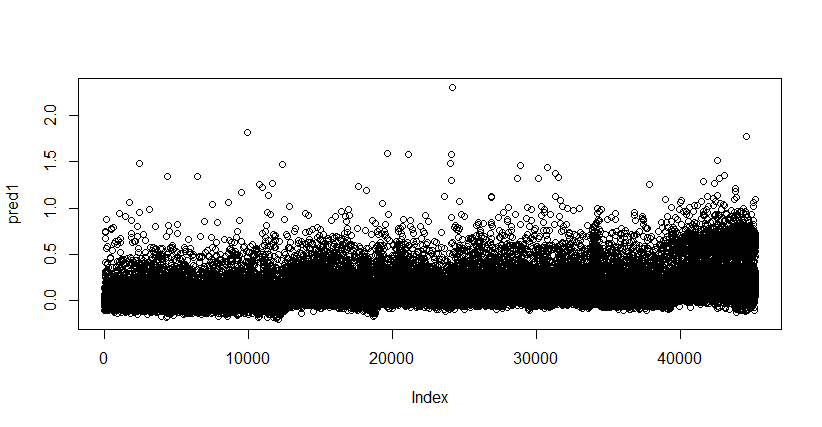
################***logistic regression***#################

**BUSSINESS PROBLEM:**

Whether the client has subscribed a term deposit or not.

PLOT:



GLM functions uses sigmoid curve to produce desirable results.

the output of sigmoid function lies in between 0 and 1.

CREATING MODEL:

Create a model with all the dependent variables and check the p value and remove the dependent variables with the highest p value and create a model.

Check the difference between null deviance and residual deviance

If the difference between the null and residual is more then the model is good.

Check the model AIC value also, if the change in AIC value is <=2 then we need to use the model with the simplest form with the low inputs.

Model1: model created with all the dependent variables

model <- glm (y ~., data = bank\_data,family = 'binomial')

Null deviance: 32631 on 45210 degrees of freedom

Residual deviance: 22640 on 45183 degrees of freedom

AIC: 22696

It consists of p values which are more than the given significant values .

So, remove the variables with the higher value and rebuild the model with the required variables with significant p value.

Model 10:

model10 <-glm (y~balance+housing+loan+duration+campaign+poutfailure+poutother+poutsuccess+con\_cellular+con\_telephone+divorced+married+joadmin.+joblue.collar +johousemaid+jomanagement+joretired+jostudent,data = bank\_data,family = 'binomial')

# Null deviance: 32631 on 45210 degrees of freedom

#Residual deviance: 22648 on 45192 degrees of freedom

#AIC: 22686

Null deviance describes goodness of fit of GUM model and discrepancy in the model

Difference in the AIC value >2 so the we go with model we built.

CONFUSION MATRIX:

0 1

FALSE 39017 3581

TRUE 905 1708

Accuracy

0.9007764

CALCULATE THE BELOW MATRTIX

0 1

0 39017 905

1. 3581 1708

precision

0.915935

TPR

0.9773308

TNR

0.3229344

FP\_rate

0.6770656

FN\_rate

0.0226692

F1

0.9456374

F1 value is nearer to 1 so the model build is a balanced model

ROC(Receiver Operating Characteristic):

More the area under the curve better the model is.

